A SINGLE CASE STUDY OF ARTICULATORY ADAPTATION DURING ACOUSTIC MIMICRY

Eleanor Lawson[a], *James M. Scobbie*[a] *& Jane Stuart-Smith*[b]

[a]CASL Research Centre, Queen Margaret University, UK; [b]University of Glasgow, UK
elawson@qmu.ac.uk; jscobbie@qmu.ac.uk; Jane.Stuart-Smith@glasgow.ac.uk

ABSTRACT

The distribution of fine-grained phonetic variation can be observed in the speech of members of well-defined social groups. It is evident that such variation must somehow be able to propagate through a speech community from speaker to hearer. However, technological barriers have meant that close and direct study of the articulatory links of this speaker-hearer chain has not, to date, been possible. We present the results of a single-case study using an ultrasound-based method to investigate temporal and configurational lingual adaptation during mimicry. Our study focuses on allophonic variants of postvocalic /r/ found in speech from Central Scotland. Our results show that our informant was able to adjust tongue gesture timing towards that of the stimulus, but did not alter tongue configuration.

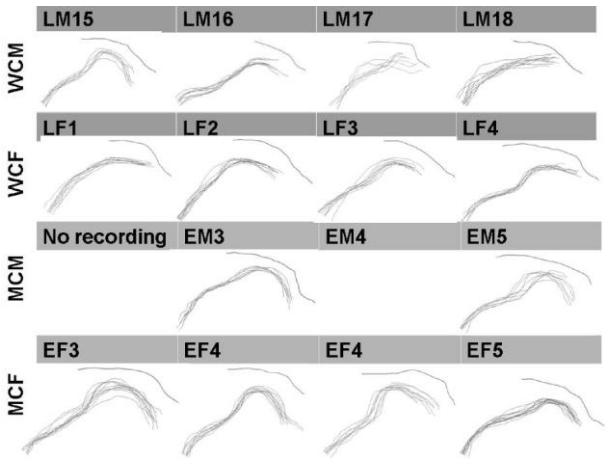
Keywords: ultrasound, mimicry, articulatory phonetics, postvocalic /r/

1. INTRODUCTION

Sociolinguistic studies of speech in Central Scotland from the late 1970s to the present day, e.g. [9, 10, 11] have identified a continuum of auditory allophonic variants of coda /r/ with a social-indexical function. At one end of the continuum are auditorily *strong* forms of /r/ [kʰaɻ] "car", used by middle-class (MC) speakers, at the other end, /r/s are weakly audible, often with accompanying pharyngealisation of the prerhotic vowel e.g. [kʰaˤ], in working-class (WC) speech. These latter variants are subsequently referred to as "derhoticised" variants.

Recent research using ultrasound tongue imaging (UTI) corpora, collected in Central Scotland, has shown that underlying these markedly different auditory variants are differences in both anterior lingual gesture-timing and tongue-configuration. For example, a study of Scottish postvocalic /r/ identified that in CVr## words such as *car, fur, beer*, where there were no anticipatory coarticulatory pressures associated with following coronals, WC informants were more likely to use tongue-tip/front raised variants, while MC informants were more likely to use bunched variants [4] (see Fig. 1).

Figure 1: All tongue surface splines (between 9 and 12 splines per informant) from CVr words in an ultrasound corpus, organized by socioeconomic and gender group. The tongue-surface contours of WC speakers occupy the top two rows and MC speakers, the bottom two rows. Above each set of tongue splines is a hard-palate trace.



Derhoticised /r/ variants were also observed to exhibit a temporal lag, whereby the constriction for /r/ was not more open, but rather temporally delayed beyond the offset of voicing, rendering some or all of the /r/ articulation inaudible, [5]. In comparison, the temporal point of maximum constriction of the anterior gesture in bunched /r/ (i.e. involving the tongue dorsum and palate), usually occurred closer to the syllable centre, well before the offset of voicing.

The fact that postvocalic /r/ variants exhibiting temporal/configurational differences are used by well-defined social groups in Central Scotland, suggests that these aspects of articulatory variation can propagate through a speech community.

We devised a method for investigating speaker adaptation towards these types of variation during mimicry, making use of pre-collected UTI-audio

corpora. Below we present the method and results of a single-case study in speaker-hearer articulatory adaptation during mimicry.

1.1. Aim

To quantify temporal/configurational adaptation of lingual articulation during mimicry, using baseline and mimicked UTI video recordings.

1.2. The informant

We used a single informant, henceforth "informant 1", aged 32 and originally from western Central Scotland, having also lived in eastern Central Scotland for 10 years. Informant 1 had been identified impressionistically by the researchers as a variably derhoticising speaker and had identified himself as having a "variable accent"; speaking one way with friends, but adopting his "lecturer's voice" in professional situations.

1.3. Experimental stimuli

Audio stimuli (all of which were high-frequency words) were extracted from audio-UTI corpora, recorded in a sound-proofed recording studio during 2007-8. In addition to various ad hoc recordings of speakers, male and female, aged 21+, mainly from Central Scotland, materials were drawn from the corpus ECB08, containing examples of speech from 12-13 year old male and female informants from the eastern Central Scotland. 24 audio recordings of words containing /r/ were extracted from the word list and sponteneous speech sections of the corpora. Only CVr##, or CVrC words were chosen, where the closing consonant was a plosive, to facilitate articulatory analysis of gesture timing in relation to the offset of voicing. Approximately equal numbers of audio files associated with /r/ articulations were chosen from three predetermined categories identified by the first author [4]: *tip-up approximant*, *bunched approximant*, and *derhoticised* (i.e. a tip-up approximant with gestural delay in relation to voicing offset). We also used 13 distractors, to shift the emphasis of the experiment away from postvocalic /r/. In total 40 acoustic stimuli were presented to informant 1.

1.4. The experimental setup

Informant 1 wore an aluminium stabilising headset to hold the ultrasound probe in place under the chin and minimise probe rotation and lateral movement, while allowing the informant to move their head, body and arms, [6].

A baseline set of audio recordings were obtained using *Articulate Assistant Advanced* (AAA) UTI software [12].

The informant was positioned in front of a monitor and prompted to produce each word using orthographic prompts presented on the monitor, while his audio and lingual movements were recorded. Immediately after recording the baseline articulations, we recorded the informant producing a set of mimicked utterances. For the mimicked recording block, the informant was asked to listen to audio stimuli which would be presented to him once only via a pair of headphones and mimic the speaker's pronunciation when the monitor screen changed colour.

2. ANALYSIS

In comparison with other studies investigating imitative fidelity, e.g. [2, 8], the innovation in this study is that we can directly compare the timing of a speaker's articulatory gestures and the speaker's midsagittal tongue shape in baseline and mimicking conditions.

2.1. Tongue gesture timing

The first author annotated the temporal point of maximum constriction for /r/ (rmax) and the onset (v-on) and offset (v-off) of voicing for each baseline and mimicked token. For informant 1, *rmax* was always the temporal point where the tongue tip was raised highest. *V-on* and *v-off* were annotated with reference to a spectrographic recording of the acoustic signal. The temporal difference between *rmax* and *v-off* was then calculated. A positive value indicated that rmax occurred after the offset of voicing and a negative value if rmax occurred before the offset of voicing. This value was then recalculated as a proportion of the voiced section of the syllable in order to take into account variation in speed of pronunciation.

2.2. Tongue configuration

Baseline refers to data obtained from UTI recordings created using orthographic stimuli only. These recordings were obtained prior to the mimicking block. *Mimicked* refers to the UTI recordings obtained using audio stimuli only and where the speaker was attempting to mimic audio recordings.

A spline was fitted to the midsagittal tongue surface in the UTI video frame closest to rmax (see Fig. 2).

Figure 1: (Left) UTI video frame before spline fitting. (Right) UTI still frame after spline fitting – informant 1's /r/ in a mimicked token of *bore*.

An average tongue surface contour was created using *mimicked-bunched* tokens from the *mimicked* set, and then, using the same lexical items, from the *baseline* set, for the purpose of comparison. Only the *baseline* and *mimicked-bunched* tongue contours are presented here, as informant 1's baseline tongue configuration is similar to that of the tip-up and derhoticised stimuli.

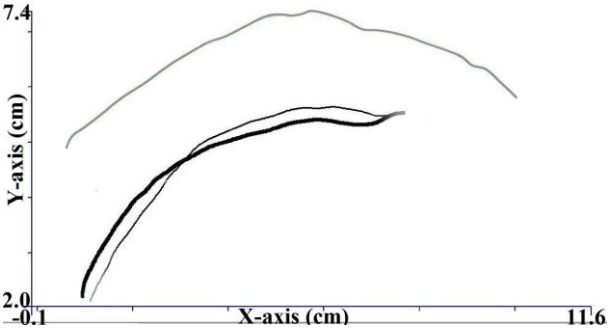
Informant 1 had difficulty mimicking audio tokens extracted from the spontaneous speech section of the corpus, even though care had been taken to avoid rapid or slurred audio examples during the selection process. This finding shows that coarticulatory cues present in connected speech can be misinterpreted when a word is taken from its phonetic context and played in isolation. Spontaneous speech recordings were therefore not analysed.

3. RESULTS

3.1. Tongue configuration

Fig. 3 below compares the average midsagittal tongue surface contour in baseline and mimicked-bunched condition, obtained using tokens of the words: verb, for, sure, bear, par.

Figure 3: Average tongue configurations for 5 baseline (grey) and 5 mimicked *bunched* tokens (black). The uppermost line in the figure shows the surface of informant 1's palate.



The informant, when mimicking a stimulus associated with a bunched tongue shape, showed little deviation from his baseline configuration. There is no adjustment towards the kind of configuration shown in the lower two panels of Fig. 1. (i.e. with a raised tongue middle and lowered tongue tip). At first sight, it would appear that the average mimicked-bunched spline shows subtle root retraction and tongue-tip lowering in comparison with the baseline spline; however, analysis of palate traces obtained at the beginning of the *baseline* and *mimicked* recording blocks, showed that slight tilting of the probe had occurred between these two sets. We therefore conclude that informant 1 made no adjustment of his tongue configuration while mimicking the bunched stimuli included in the analysis. For the other mimicked tokens (tip-up and derhoticised), informant 1's mean tongue shape was also almost identical to that of the mean baseline. Informant 1 produced a front-bunched tongue shape on one occasion when mimicking one of the excluded spontaneous-speech tokens containing a bunched /r/, *start* [stɑɻʔ], which he mimicked as a nonword, [dʌːː].

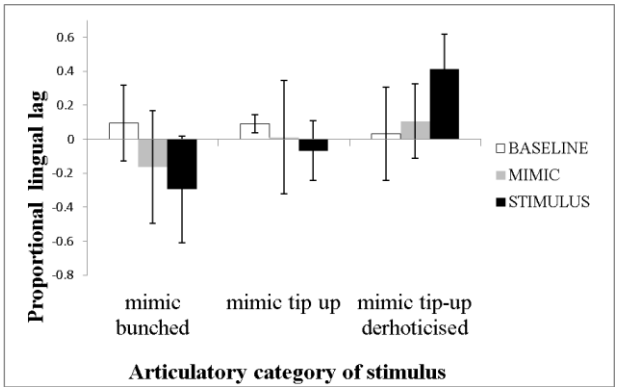
3.2. Tongue gesture timing

Analysis of gesture timing does evidence accommodation towards the stimulus articulations. Fig. 4 below compares informant 1's mean *baseline* and *mimicked* proportional gesture lag to that of the stimuli. The white and grey bars show informant 1's proportional gesture lag in baseline and mimicked condition respectively and the black bars show the mean proportional gesture lag of the stimuli. The data are split into three different tongue configuration types (*tip-up*, *bunched*, and *derhoticised*), which were previously observed to exhibit different rmax timings in relation to the offset of voicing. *Bunched approximant* variants tend to have an early point of maximum constriction. *Tip up approximant* variants tend to have a point of maximum constriction close to the offset of voicing, and *derhoticised* /r/ variants tend to have a point of maximum constriction after the offset of voicing.

Informant 1's baseline recordings all show a slight positive lag. In the mimicked condition, he alters his *rmax* timings towards those of the stimuli. Informant 1's adaptation is greatest in the *mimic-bunched* condition where his baseline *rmax* lag is adjusted from +10% of the duration of the voiced section of the syllable to -16%. He adjusts his

baseline lag from +9% to to +1% in the *mimic tip-up* condition and he adjusts his baseline lag from +3% to +10% in the *mimic-derhoticised* condition.

Figure 4: A comparison of the proportional gesture lag in the stimuli and informant 1's baseline and mimicked recordings. Whiskers represent +/- one standard deviation.



4. DISCUSSION

When mimicking acoustic input, informant 1 adapted *rmax* timing, but not the configuration of his tongue. The informant's lack of adjustment of tongue configuration during mimicry may support the notion of articulatory tradeoffs [1, 3], where radically different tongue configurations can be used by speakers to produce similar acoustic outputs. However, it is possible that closer tongue-configuration adaptation can occur where there is no lexical access, as in the example of mimicked spontaneous speech detailed above, see also [2].

Informant 1's adjustment of gesture timing shows that adaptation towards subtle timing variation is possible. In a shadowing study of voicing timing, Mitterer and Ernestus [7] found that only phonological relevance controlled imitative tendencies. However, the ability to perceive and imitate subphonemic variation in this single-case study is unsurprising, given the indexical function of fine-grain phonetic variation in Central Scottish postvocalic /r/.

It is unclear whether informant 1's imitative behavior is typical of all speakers. It is possible that speakers from different gender, age and social-class groups may exhibit different imitative strategies. A more comprehensive study is needed in order to find out if informant 1's responses are typical, or if closer imitation of tongue configuration is likely when different types of stimuli, e.g. nonsense words, are used. What seems clear, even at this stage, is that the acoustic quality of /r/, which is socially variable, is due to an interplay of shape and timing.

5. REFERENCES

[1] Alwan, A, Narayanan, S. 1996. Towards articulatory-acoustic models for liquid approximants based on MRI and EPG data. Part II. The rhotics. *JASA* 101(2), 1078-1989.

[2] Goldinger, S.D. 1998. Echoes of echoes? An episodic theory of lexical access. *Psychological Review* 105, 251-279.

[3] Guenther, F.H., Espy-Wilson, C.Y., Boyce, S.E., Matthies, M.L., Zandipour, M., Perkell, J.S. 1999. Articulatory tradeoffs reduce acoustic variability during American English /r/ production. *JASA* 105, 2854-2865.

[4] Lawson, E., Scobbie, J.M., Stuart-Smith, J. 2011. The social stratification of tongue shape for postvocalic /r/ in Scottish English. *J. of Sociolinguistics* 15, 256-268.

[5] Lawson, E., Stuart-Smith, J., Scobbie, J.M. 2008. Articulatory insights into language variation and change: Preliminary findings from an ultrasound study of Derhoticization in Scottish English. *Selected Papers from NWAV 35, University of Pennsylvania working Papers in Linguistics* 14, 102-109.

[6] McLeod, S., Wrench, A.A. 2008. Protocol for restricting head movement when recording ultrasound images of speech. *Asia Pacific Journal of Speech, Language and Hearing* 11, 23-29.

[7] Mitterer, H., Ernestus, M. 2008. The link between speech perception and production is phonological and abstract: Evidence from the shadowing task. *Cognition* 109, 168-173.

[8] Nye, P., Fowler, C.A. 2003. Shadowing latency and imitation: The effect of familiarity with the phonetic patterning of English. *Journal of Phonetics* 31, 63-79.

[9] Romaine, S. 1979. Postvocalic /r/ in Scottish English: Sound change in progress? In Trudgill, P. (ed.), *Sociolinguistic Patterns in British English*. London: Edward Arnold, 145-157.

[10] Speitel, H., Johnston, P., 1983. *A Sociolinguistic Investigation of Edinburgh Speech.* ESRC End of Grant Report.

[11] Stuart-Smith, J., 2007. A sociophonetic investigation of postvocalic /r/ in Glaswegian adolescents. *Proceedings of the XVIth ICPhS* Saarbrücken, 1449-1452.

[12] Wrench, A.A., 2007. *Articulate Assistant Advanced User Guide* (Version 2.07). Edinburgh: Articulate Instruments Ltd.